Task-Specific Compression for Biomedical Big Data

UNIVERSITY OF ARIZONA

PI: BILGIN, ALI   Grant Number: 1 U01 CA198945-01

Contemporary biomedical research increasingly generates and uses very large datasets. As this seemingly unending supply of biomedical Big Data is collected, processed, and stored, a key challenge is maintaining and delivering this data efficiently. The challenges associated with Big Data are particularly prominent in digital pathology: A single digital whole slide image (WSI) requires a file size in the 2 to 10GB range. When an entire case is considered (typically between 4 to 30 different stains), the raw data size can exceed 100 GB. With imaging at different focal planes (z-stacks) and multispectral imaging becoming available, it is not unreasonable to expect that the raw data from a single case will reach several TBs in the near future. The slide volumes of a typical academic pathology department require round-the-clock operation of multiple scanners which can be loaded with hundreds of slides and can scan continuously. Thus, the volume of data that is expected to be generated by a fully digital pathology practice is enormous. The goal of this proposal is to solve this challenging problem. Our hypothesis is that it is possible to significantly improve the presentation of digital pathology images for accurate diagnoses by designing intelligent image compression schemes. We propose three Specific Aims to test this hypothesis: Aim 1: To develop and validate efficient and intelligent image compression techniques optimized based on the properties of the Human Visual System (HVS). These techniques will optimally tune the compression parameters such that desired visual quality is achieved for each and every image. Aim 2: To develop and validate a novel image compression paradigm where information which is most relevant to the task at hand is stored and transmitted preferentially. We propose to use Task-Specific Information (TSI) as a metric of image fidelity during compression. Aim 3: To develop a client-server framework that will allow interactive remote browsing of very high resolution pathology images over bandwidth-limited networks.   PUBLIC HEALTH RELEVANCE PUBLIC HEALTH RELEVANCE: Digital imaging technologies provide unprecedented opportunities in medicine. However, enormous data sizes created by today's high resolution medical imaging techniques remain as a significant technological barrier. In digital pathology, the raw data for a single case can exceed 100 GB without compression. The major hypothesis of this project is that it is possible to design image compression schemes that will improve the presentation of digital pathology images for accurate diagnoses, and that such intelligent compression schemes will yield significant compression ratios.